



Application des techniques de Machine Learning à la géolocalisation indoor

Abdelilah Mejdoubi

Plan

- **Technologies, techniques et applications de localisation indoor**
- **Caractéristiques de la base UJIIndoorLoc**
- **Data Preprocessing**
- **Prédiction**
- **Conclusion**

Applications de localisation indoor

- Intelligence ambiante (objets à localiser les uns par rapport aux autres)
- Applications de guidage de personnes (milieu hospitalier, musée, automobile, etc.)
- Surveillance de personnes à risques (prisonniers, enfants en bas âge, etc.)
- Dans les bâtiments (Gestions d'alarmes, obtention de statistiques sur des parcours)



Techniques de localisation indoor

- Les dispositifs RFID, La localisation par identification de cellule : Cell ID, La localisation par mesures temporelles (TOA/TDOA), etc
- La localisation par mesure de puissance de signal (fingerprinting / RSSI)
 - Information simple à mesurer par les interfaces radio
 - Tous les équipements WiFi sont donc "localisables" sans modification

Exploitation de la technique du fingerprinting pour se localiser

- Besoin de créer une base de données
- Localisation : recherche dans cette BDD de celle qui est la plus proche de la mesure instantanée

Inconvénients

- Besoin de constituer une carte d'empreintes des puissances par enregistrement en plusieurs points de l'espace (calibrage)
- Carte "invalidé" si l'environnement change

Caractéristique de la base UJIIndoorLoc

- Couvre une surface de 108 703 m² - 3 bâtiments de 4 ou 5 étages
- 19737 échantillon d'apprentissage et 1111 échantillon de test/validation – 21048 au total
- 520 différents WAPs (Wireless Access Points)
- L'intensité des signaux varie de -104 dBm (faible intensité) à 0 dbm (forte intensité).
La valeur 100 dBm est utilisé pour indiquer qu'il n'y a pas de signal détecté
- Les données sont collectées par 18 utilisateurs avec 25 modèles d'appareils mobiles différents
- Les coordonnées latitude, longitude, étage et l'ID du bâtiment sont fournis comme des attributs à prédire.

Data Preprocessing

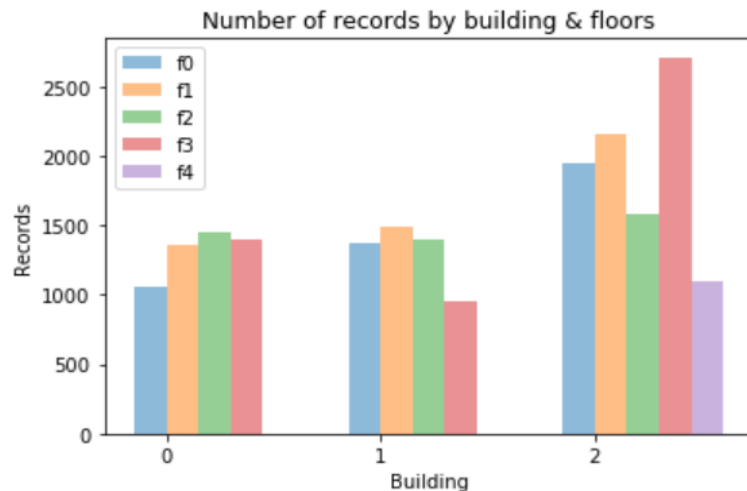
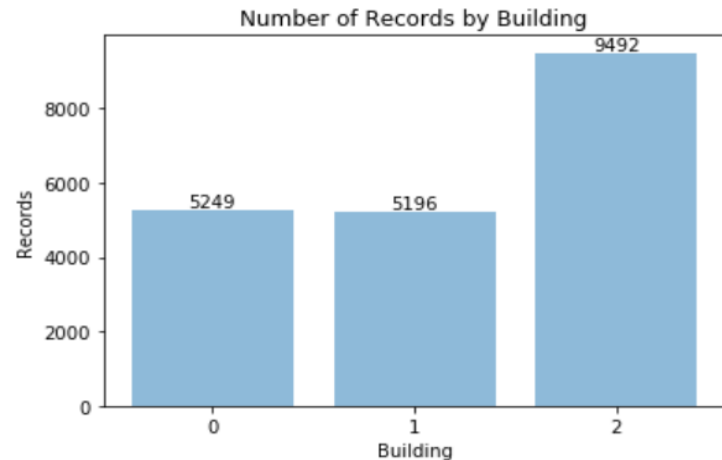
- Aucune donnée manquante pour le training et le test

Nombre d'enregistrements par bâtiment & étage

Seul le bâtiment 2 a le niveau 5

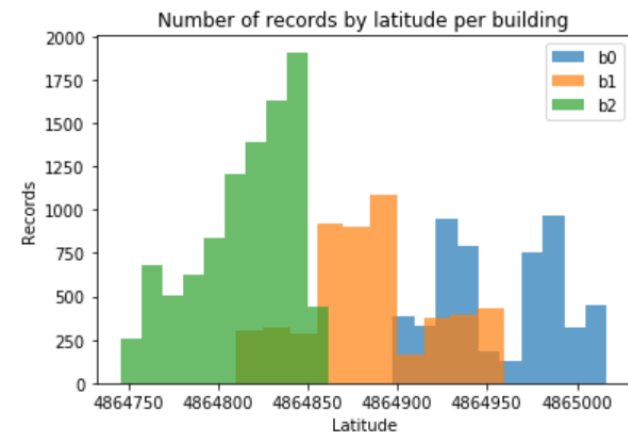
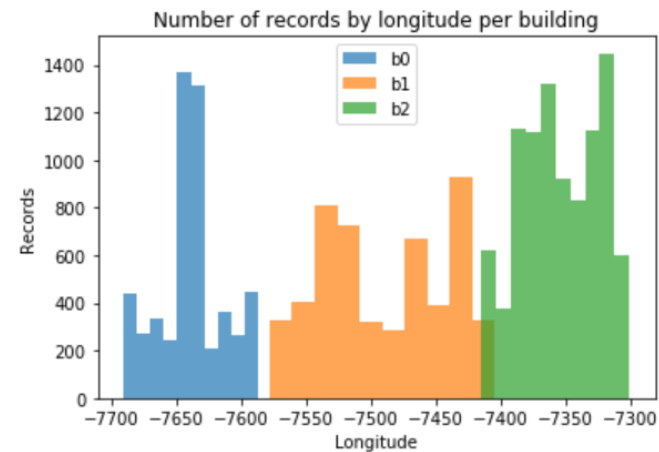
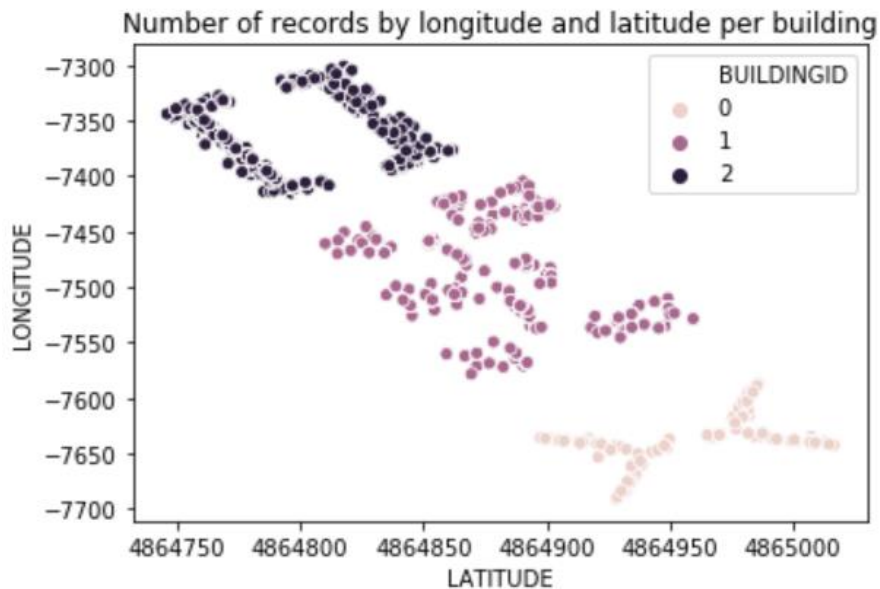
Répartition relativement uniforme sur tous les étages des bâtiments 0 et 1

Nombres d'enregistrements plus élevé au niveau 4 pour le bâtiment 2



Data Preprocessing

Nombre d'enregistrements par Longitude, Latitude & bâtiment



Data Preprocessing

Réduction de dimensions

- 520 colonnes WAP: pour chaque enregistrement, tous les variables d'entrées sont remplacées dans une seule colonne [MAXSIGNAL] par la puissance du signal la plus élevée sur tous les WAP
- Certain enregistrements n'ont pas de valeur du signal (pas de lecture RSSI)

MAXSIGNAL

-53.0

-46.0

-61.0

-55.0

NaN

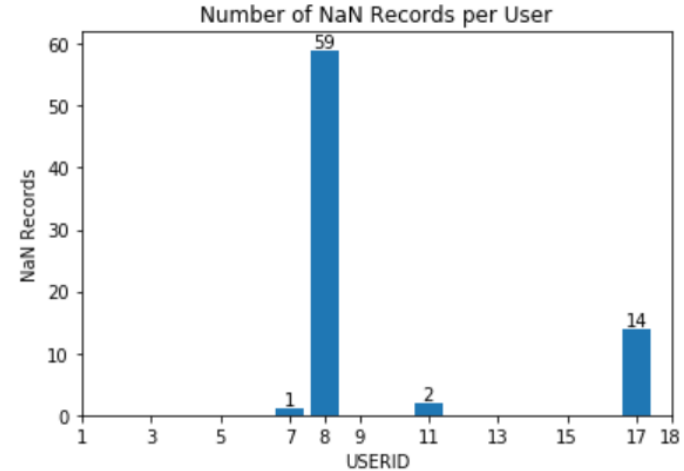
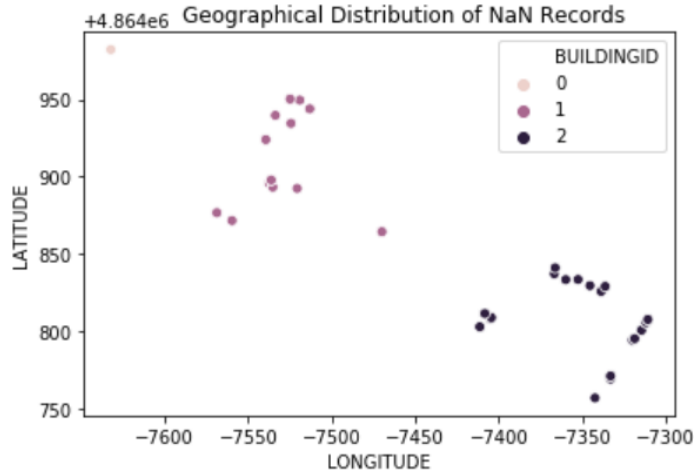
Data Preprocessing

Analyse des enregistrements des signaux non détecté

- Nombres d'enregistrements des signaux non détecté : 76 (0,38%)
- La majorité des enregistrements NaN proviennent de l'utilisateur 8 : 59 signaux (77,63%)
- Nombres d'enregistrements NaN dans la base VALIDATION : 0 (0%)

	USERID	RECORDS
0	7	1.32%
1	8	77.63%
2	11	2.63%
3	17	18.42%

55 enregistrements NaN sont supprimer de la base d'apprentissage



Prédiction : bâtiment & Etage

Algorithme de classification

- 520 variables d'entrées (dont 55 variables sont exclus)
- Variables cibles: bâtiments et étages
- 13 classes

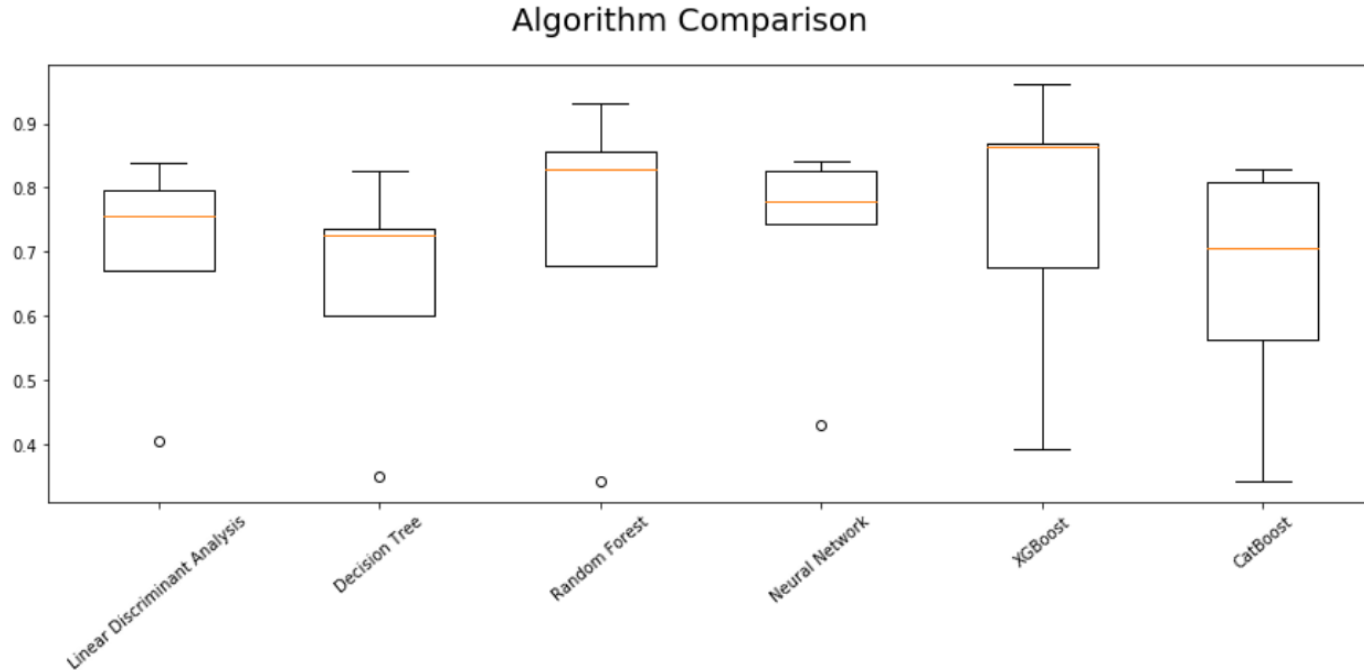
Modèles utilisés

- Linear Discriminant Analysis
- Decision Tree
- Random Forest
- Neural Network
- XGBoost

- Pour estimer les performances de chaque modèle, on a utilisé 5-fold cross validation

Prédiction : bâtiment & Etage

Algorithme de classification - Modélisation sans paramètres



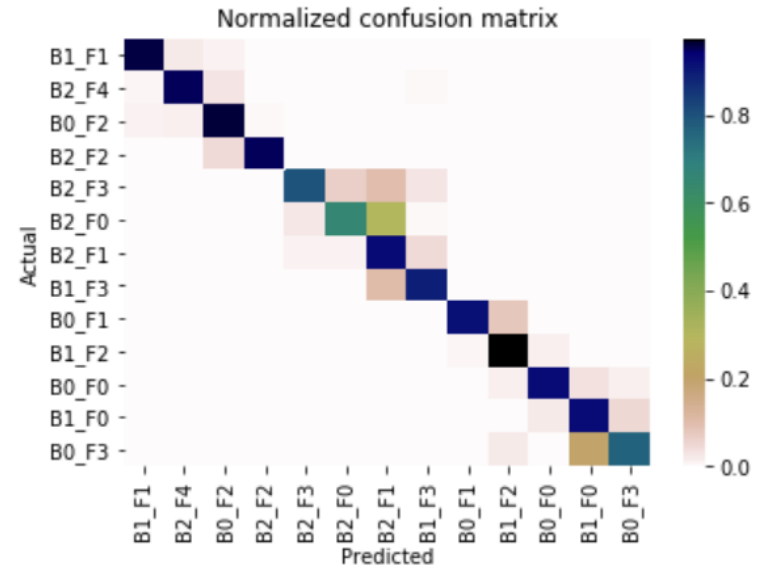
XGBoost et Random Forest donnent une meilleure performance

Prédiction : bâtiment & Etage

Algorithme de classification - XGBoost (Modélisation avec paramètres)

Accuracy = 87,67 %

```
"learning_rate":0.1,  
"n_estimators":100,  
"max_depth":5,  
"min_child_weight":1,  
"gamma":0,  
"alpha":0,  
"lambda":1,
```



Prédiction : bâtiment & Etage

Algorithme de classification - XGBoost (RandomizedSearchCV) & Random Forest

Utilisation de l'algorithme PCA : 200 variables d'entrée

Best paramètres

```
{'gamma': 0,  
 'learning_rate': 0.27117147016034093,  
 'max_depth': 4,  
 'min_child_weight': 3,  
 'n_estimators': 390,  
 'objective': 'multi:softmax'}
```

XGBOOST

Base accuracy: 87.67%
Best accuracy from Randomized Search: 91.25%
Improvement of 4.09%

Comparaison

	XGBOOST improvements	Random Forest
Buildings	99.91 %	99.73 %
Floors	90.10 %	87.31 %

Prédiction : Latitude & Longitude

Algorithme de régression

- 520 variables d'entrées (dont 55 variables sont exclus)
- Variables cibles : Latitude et Longitude

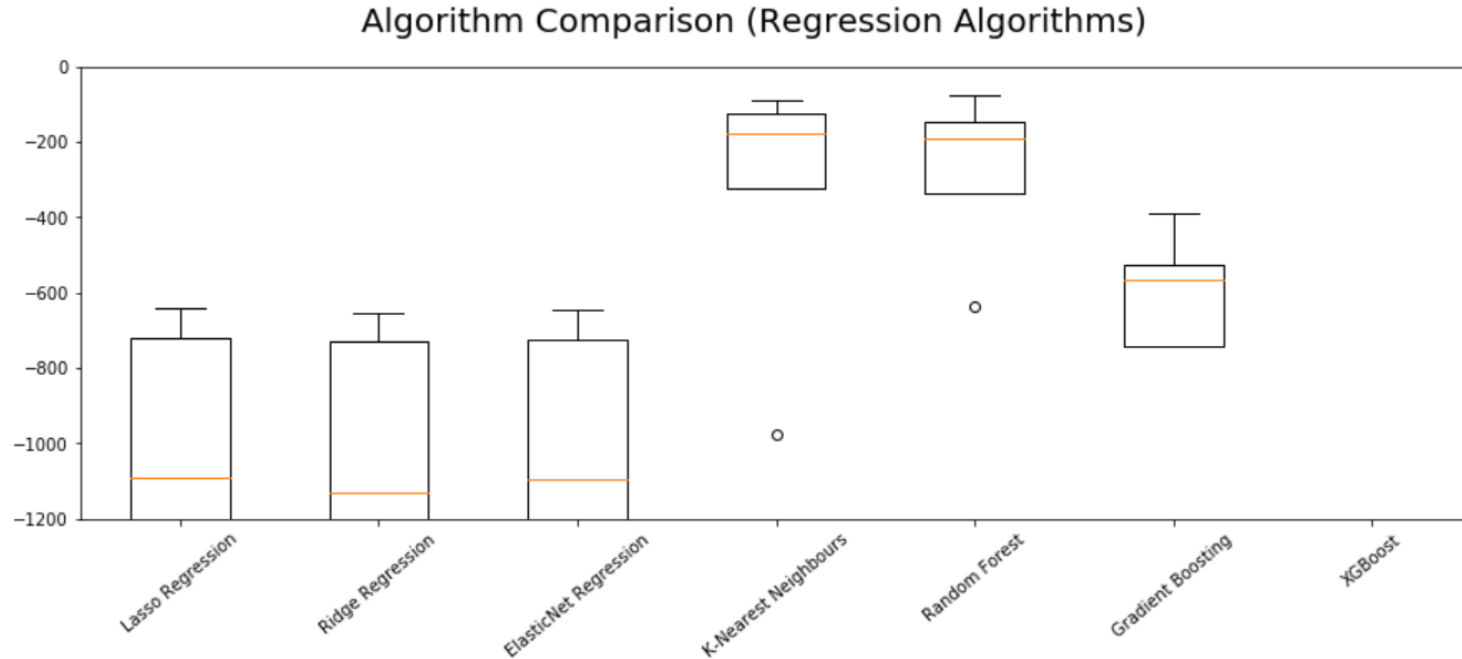
Modèles utilisés

- Lasso Regression
- Ridge Regression
- ElasticNet Regression
- K-Nearest Neighbours
- Random Forest
- Gradient Boosting
- XGBoost

- Pour estimer les performances de chaque modèle, on a utilisé 5-fold cross validation

Prédiction : Latitude & Longitude

Algorithme de régression - Modélisation sans paramètres



La valeur du Mean Square Erreur est plus faible pour les modèles KNN et Random Forest

Prédiction : Latitude & Longitude

Algorithme de régression - K-Nearest Neighbours & Random Forest

K-Nearest Neighbours

R-squared score: 0.9696

Mean Squared Error:

Latitude: 206.76

Longitude: 272.66

Random Forest

R-squared score: 0.9739

Mean Squared Error:

Latitude: 177.23

Longitude: 236.41

En moyenne :

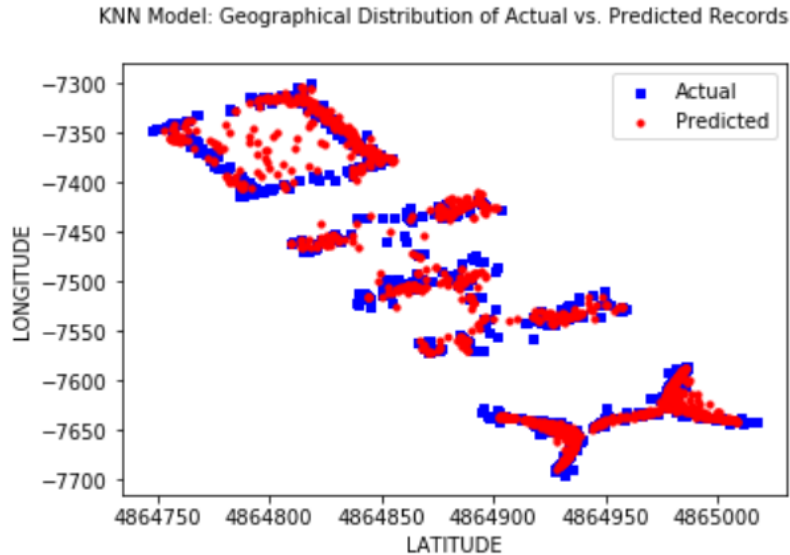
- Le modèle KNN est précis jusqu'à un rayon de 13,15 m
- Le modèle Random Forest est précis jusqu'à un rayon de 14,28 m

La distance entre la valeur actuel et la valeur prédite est calculé en utilisant le théorème de pythagore

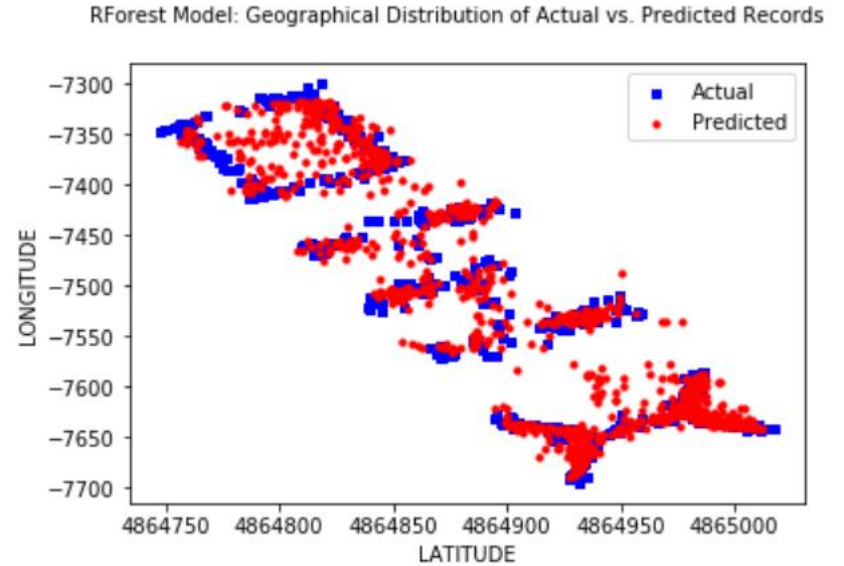
Prédiction : Latitude & Longitude

Algorithme de régression - K-Nearest Neighbours & Random Forest

K-Nearest Neighbours



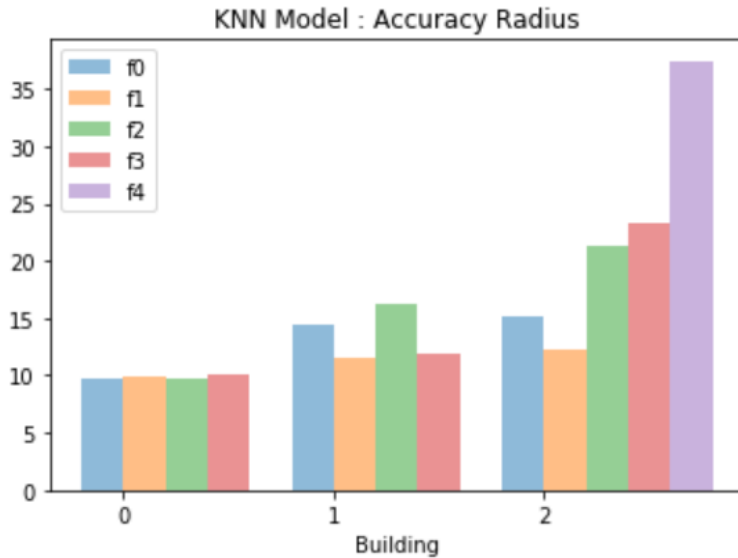
Random Forest



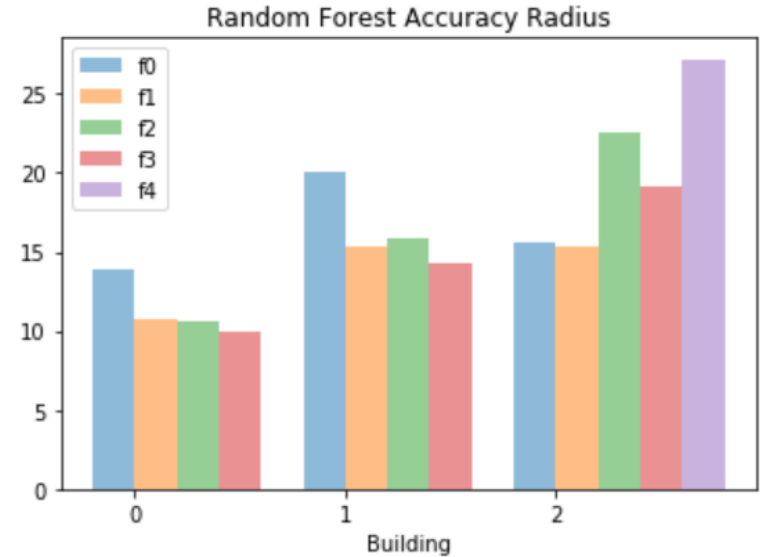
Prédiction : Latitude & Longitude

Algorithme de régression - K-Nearest Neighbours & Random Forest

KNN

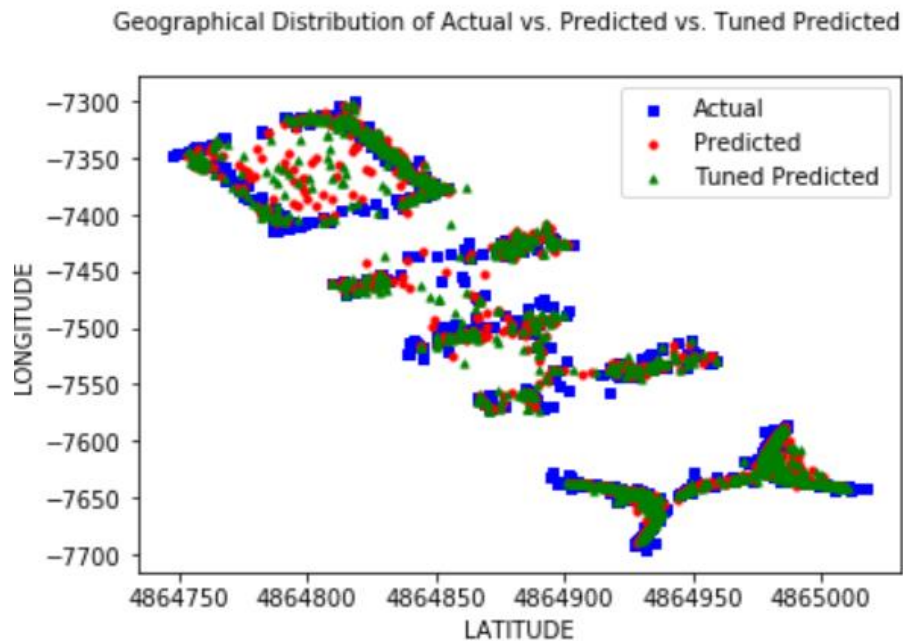
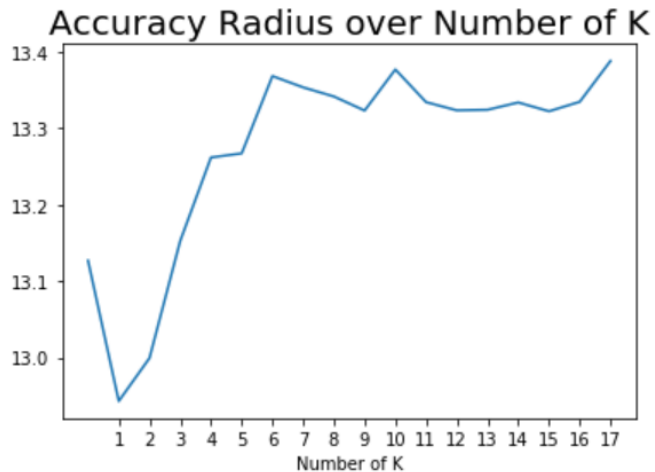


Random Forest



Prédiction : Latitude & Longitude

Algorithme de régression - K-Nearest Neighbours – Amélioration du modèle



Conclusion

- Les modèles XGBoost, Random Forest et KNN donnent une meilleure performance
- La prédiction de la position pour les bâtiments est plus élevée que celle de l'étage

Amélioration

- Ajouter d'autres dimensions dans les modèles, par exemple examiner l'effet du temps sur les prévisions



mejd.abdel@gmail.com

06 10 05 36 17

Questions ?